We divided our project into three parts according to the assignment1.

The code provided is a Hadoop MapReduce software designed to find mutual friends in a dataset. This identifies mutual friendships between individuals. The code has two main parts:

MyMapper and MyReducer.

Task 1: Deepika Murakonda

In the MyMapper, information from the dataset is read and analyzed. Each line of dataset is first split by space identifying persons and their friends. Once there are split by persons and friends, it then performs the pairing of friends in a way that they are sorted. For every key [pair of friends] their values [ mutual friends] were assigned. All pairs are grouped by using their keys. The output from the mapper would be the pair of friends as key and they’re all possible friends related to that pair as value. This output will be further reduced by the reducer.

The MyReducer receives intermediate key-value pairs generated through the mapper. The reducer identifies all possible mutual friends from all the related pairs. We simultaneously calculated the sum of mutual friends and count of the mutual such that we can calculate the average number of mutual friends and the maximum number of the mutual friends. The reducer results an output as a key value pair, where key is the pair of friends and value will be the mutual friends of the key. Here we did not consider those keys who has no mutual friends.

Task2: Nikhil Vuppala

Now we have a context which contains the keys as pair of friends and the list of mutual friends between them.

After processing all key-value pairs, I have done small modification in the Task1 which is nothing but finding the Highest number of mutual friends in which the variable is assigned outside the Reduce method and we are taking a HashMap outside the reduce method which we store the key and value for the further tasks.

Now we use the concept of a cleanup method in which this method calls by default after the reduce method. The variable which we stored in the My reducer can access in the cleanup method. So, by iterating the HashMap we print the key and values which are equal to the highest number of mutual friends

Now for sub question the values of the keys whose values are starting with “1” or “5”. We are printing only those values with the keys by iterating the hash and again iterating the list of mutual friends.

Task3: Sahithi Thota

To calculate the average of the mutual friends of the data set. I am assigning two variables outside the reduce map in which I have done a slight modification in task1 and again removing the empty mutual friends which is not necessary by using the if condition and increment in the total count of the mutual friends. So, that overall average will be maximum.

With the help of the overall average, we need to iterate the HashMap once again and print the list of mutual friends whose length is more than the overall average value.

Note: All the three tasks are in one output file, which we easy to differentiate by searching the Task2, Task3 by that it goes to their respective line.

A screenshot of a test

Description automatically generatedTask1 Output:

Task2 Output:

A close-up of a number

Description automatically generated

Task3 Output:

A close-up of a number

Description automatically generated